SERI/TP-351-545
UC CATEGORY: UC-60
CONF-800804--18

PLANNING FOR ELECTRIC UTILITY SOLAR APPLICATIONS: THE EFFECTS ON RELIABILITY AND PRODUCTION COST ESTIMATES OF THE VARIABILITY IN DEMAND

GEORGE R. FEGAN
C. DAVID PERCIVAL

JANUARY 1980

TO BE PRESENTED AT THE CENTURY 2 SOLAR ENERGY CONFERENCE, SAN FRANCISCO, CALIFORNIA, AUGUST 1980

PREPARED UNDER TASK NO. 3532.20

Solar Energy Research Institute

1536 Cole Boulevard
Golden, Colorado 80401

A Division of Midwest Research Institute

Prepared for the
U.S. Department of Energy
Contract No. EG-77-C-01-4042

DISCLAIMER

MASTER

SERI/TP-351-545
UC CATEGORY: UC-60

PLANNING FOR ELECTRIC UTILITY SOLAR APPLICATIONS: THE EFFECTS ON RELIABILITY AND PRODUCTION COST ESTIMATES OF THE VARIABILITY IN DEMAND

GEORGE R. FEGAN
C. DAVID PERCIVAL

JANUARY 1980

TO BE PRESENTED AT THE CENTURY 2 SOLAR ENERGY CONFERENCE, SAN FRANCISCO, CALIFORNIA, AUGUST 1980

PREPARED UNDER TASK NO. 3532.20

Solar Energy Research Institute

1536 Cole Boulevard
Golden, Colorado 80401

A Division of Midwest Research Institute

Prepared for the
U.S. Department of Energy
Contract No. EG-77-C-01-4042

PLANNING FOR ELECTRIC UTILITY SOLAR APPLICATIONS: THE EFFECTS ON RELIABILITY AND PRODUCTION COST ESTIMATES OF THE VARIABILITY IN DEMAND

George R. Fegan C. David Percival
Solar Energy Research Institute
Golden, Colorado

Abstract - Previous studies have shown the necessity of the consideration of hourly variability in the output from the intermittent generation source. However, the studies did not take into account the variability in the demand. Major questions concerning the variabilities of demand and intermittent source output are (1) does the demand variability dwarf the intermittent output variability?; (2) can demand variability be handled in the Baleriaux-Booth framework?; and (3) what effect does the demand variability have on the LOLP criterion and the reserve margin? Before attempting answers to these questions, the term variability in demand is clarified by distinguishing between variability due to randomness and variability due to forecasting uncertainty. A result is presented which shows that under general conditions the variability due to randomness can be ignored except in the neighborhood of the peak and minimum demands. The above questions are then addressed in terms of the two types of variability in demand.

INTRODUCTION

Most of the solar applications for the electric utilities' generation system are categorized as intermittent sources; that is, their power output may fluctuate freely from zero to some maximum during small time intervals. In two previous papers [1,2], the authors have claimed that measurements of average output from intermittent sources is inadequate and that the hourly variability in output must be considered. However, there is some inconsistency in this position if one fails to take into account the variability in load or demand. In fact the most often heard objection to our position is "why should we consider the variability of the output from the solar devices when its effects are swamped by the variability of demand?"

This comment gets some theortical support from one of the most widely used formulations of the reliability/production cost problem. The basic equation for the measure of reliability in the Balereaux-Booth formulation is

$$\Pr\{[\tilde{L} - \sum_{i=1}^{N} (\mathrm{Cap}_i - \tilde{FO}_i)] > 0\}$$

where

Pr = probability

$\tilde{L}$ = load regarded as a random variable

Cap_i = capacity in MW regarded as the deterministic nameplate rating of the i-th resource not on maintenance

$\tilde{FO}_i$ = force outage in MW of the i-th source regarded as a random variable

N = the number of sources on the system

Now the conceptualization for the random variable L is usually done quite poorly as will be discussed below; the problem in conceptualization is usually due to making a transition from the concept "hours in which a certain load is exceeded" to the concept of probability. However, the observation to be made here is that the load or demand is a random variable; it possesses a nonzero variability. To adequately address the variability of the intermittent output, one must have at least an intuitive handle on the variability in demand.

Once it is recognized that this demand variability must be considered, the question is how shall it be treated. The most common tools used in the evaluation of reliability and production costs are the Calabrese LOLP calculation and the Baleriaux-Booth framework. It has been demonstrated [1] that for reliability calculations the two methods are equivalent. Therefore we will give a procedure for the incorporation of demand variability into the Baleriaux-Booth framework. Its handling in Calabrese-type calculations should follow immediately.

Since it is most common not to recognize load variability in the Baleriaux-Booth framework, the LOLP calculations based in the procedure described in this paper are quite different than those in which the variability is ignored. We will also try to establish some relationship between the LOLP calculation and the reserve margin as a percent of load in the situation in which variability in demand is being considered.

VARIABILITY DUE TO RANDOMNESS VERSUS VARIABILITY DUE TO ASSUMPTIONS

When one seeks to answer the objection that the variability in demand dominates any variability traceable to the output from the intermittent source, one must be certain as to what is meant by variability in demand. At one level, the variability reflects the randomness in demand from excursions due to weather, transitory changes in electric motor usage, and variations due to entertainment habits. To be specific a forecaster makes a prediction for energy and peak demand for the next year on a month by month basis. The forecast interval, of course, could be shorter. If in the realization of that year all the assumptions, which the forecaster made, remain true, the deviation of the actual demand from the forecast will be governed by random events. This deviation we will can random variablity.

At the next level we have variation due to assumptions. We can imagine a forecaster, faced with uncertainty in the economic sector, speculating on possible scenarios for the future, each of which would be weighted by some probability. His econometric model then will produce different demand forecasts consistent with the respective sets of assumptions. The spread of these forecasts for a given year represent what we will call variability due to assumptions or scenarios.

The answer to our mythical objector then depends upon which variability in demand is intended. Both levels of variability can be treated in a Baleriaux-Booth framework but we will argue that it is almost meaningless to handle the variability due to scenario in this fashion. Also the question of dominance is not meaningful if the error distributions of the forecast remain identical in the random variability case.

RANDOM VARIABILITY

In order to examine the random variability we must look at the very nature of $\tilde{L}$, the random variable representing load in the Baleriaux-Booth framework. Booth himself leaves this term ill-defined:

> The probability distribution of the load levels experienced by the power system may conveniently be shown in a load duration curve as that shown in Fig. 1(a). This curve relates the loading levels to the percentage of the total time that each load would be equalled or exceeded [3].

> How we get such a curve is largely immaterial to this explanation, so here we have the probablibity distribution (or density function) for the loads we are to meet [4].

If one looks at the typical method of creating a load duration curve, one gets an insight into the basis of confusion over the random variable $\tilde{L}$. Standard discussions recommend that one plot the percent of time load exceeds a particular load level versus the load level. The percent of time is then interpreted as a probability number. The problem with this is that if one is dealing with historical data and doing calculations over this history, everything is deterministic. There doesn't seem to be any probability questions which can be answered because history has already happened; there is no uncertainty involved. The conceptual trick when dealing with past history is to imagine one is at an unknown time in history; the LDC merely represents the chances of seeing loads in excess of given values. Therefore, since one doesn't know the actual load at that point in history due to the lack of chronological data representation in the LDC, one can ask questions like "what are the chances of exceeding a load of 1000 MW?" The mistake is to regard the LDC as an aggregated history over an interval; one should pretend that one is standing at a unknown instant in that time interval equipped only with information concerning a frequency distribution.

An easier way to conceive of $\tilde{L}$, the random variable for demand or load, is to look at the problem from the forecasting point of view. Let us imagine a forecaster being forced to make hourly forecasts over a given time interval. At each hour the forecaster gives an estimate of the μ (hr) = mean value for the hour. These mean forecasts give a trajectory for the load over this interval. However, the forecaster understands that as history passes and the time interval is actually realized, there are an infinite number of trajectories which could be realized. What the forecaster hopes for is that the actual trajectory will lie in a band around his forecasts of the means. Figure 1 gives a graphical representation of the process.

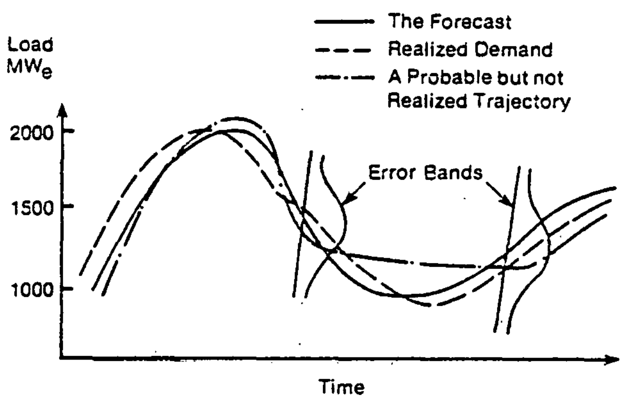


Figure 1. Mean Forecast Vs. Possible Trajectories

In reality the utility forecaster usually gives only a monthly energy and peak forecast rather than hourly forecasts; the argument is the same: he is estimating the μ(= mean) and there are an infinite number of trajectories which can be realized.

Turning our attention from the forecast to past history, one can conceptualize the seemingly deterministic event in an analogous fashion. If one had made an hourly forecast of the demand and the actual demand deviated from the forecast within reasonable limits and without systematic patterns, then one could imagine the forecast as being the mean hourly values for that year and the actual history as a single realization. If the year could be played over and over again, the hourly arithmetic averages would estimate mean values. In fact a utility system which showed little growth or change in load profile from year to year would be in a sense replaying the year. Variations in hourly values would be due to random events, one of which would be weather.

Now if one had not made a forecast for the previous years, one could use the actual demand to estimate the mean values. However, one would try to correct values which were noticeably out of bounds, for instance, a demand reaction to unusually cold weather. Of course this is what is actually done when one attempts to build a typical LDC: the utility planner either corrects for weather or averages normalized shapes over multi-years in an attempt to smooth unusual variations.

BALERIAUX-BOOTH FORMULATION FOR RANDOM VARIABILITY

Before attempting to place random variability into the Baleriaux-Booth context, one should sharpen one's concept of load duration curves (LDC's). As mentioned earlier the transition from building an LDC from realized demand values to a probabilistic interpretation causes some confusion. In Fig. 2 we see how the 8760 hourly values are normalized to give a probability scale.

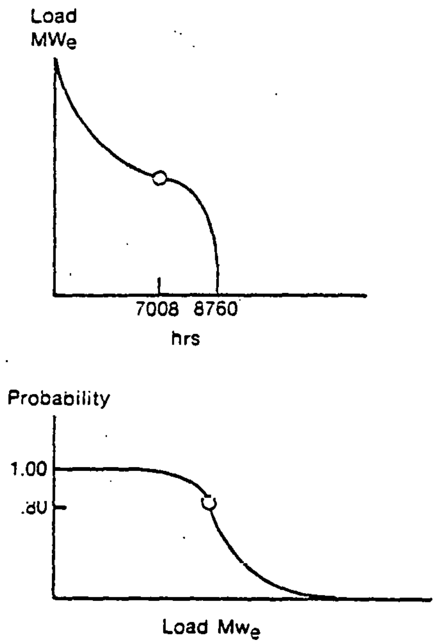


Figure 2. (a) LDC as Percent of Time (hours)
(b) LDC with Time Normalized and Axes Switched

At the point marked in the (a) figure, it is correct to say that 80 per cent of the hours exceeded this value since the LDC represents values realized. At some point, however, one must switch one's thinking from realized values to the estimate of the mean values. Therefore in the (b) figure it is not correct to say that 80 per unit of the values exceeded the marked point. Since we are now conceiving of the LDC as being constructed of mean values it is quite possible that no actual values would lie exactly on the curve; that is, the values given are expected values. In this context it is correct to say that there is an 0.80 probability that the given value will be exceeded; we can also talk about the expectation that 80 per cent of the values will exceed the given one. But to say that 80 per cent of the values will exceed this value is to conceive of the demand as a deterministic event or as a realized set of values rather than as the set of means of random variables.

It is the concept that the individual hourly means do not have to be realized which allows us to incorporate the variability due to randomness directly into the Baleriaux-Booth formulation.

The way LDC's are presently constructed is that if there are n hours in a time interval each hour receives 1/n for a probability weight. Demand values are then ranked and weights accumulated. In the present construction we propose to take the hourly forecast of the mean and the variation and distribute the 1/n weight over a range of values for the hour. These values will then be ranked, weights aggregated, and a LDC formed.

To clarify matters we will use the following simple example

Table I. (a) Hourly Demand Given as Conventional Averages
(b) Pr [Load > L]

(a)		(b)	
HR	Demand (MW)	L (MW)	Pr [Load > L]
1	100	0	1.00
2	150	100	0.50
3	100	150	0.0
4	150		

The probability values in (b) of Table II are arrived at by weighting those probability values in (a) of Table II by the probability of the hours and summing up the weights for each value, i.e. for 70 MW we have 0.20 x 1/4 + 0.20 x 1/4 where 1/4 are the weights for hours 1 and 3. The values are then ranked and the probabilities consecutively substracted from 1.

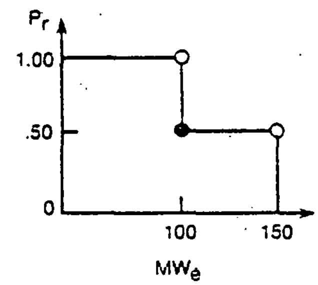
Table II. (a) Actual Forecast Range
(b) Pr [Load >L)

(a)

HR	Demand (MW)	Probability
	130	.20
	115	.20
1	μ=100	.20
	85	.20
	70	.20
	170	.35
2	μ=150	.10
	140	.40
	130	.15
	130	.20
	115	.20
3	μ=100	.20
	85	.20
	70	.20
	170	.35
4	μ=150	.10
	140	.40
	130	.15

(b)

L (MW)	Pr [Load > L]
0	1.00
70	0.90
85	0.80
100	0.70
115	0.60
130	0.425
140	0.225
150	0.175
170	0.0



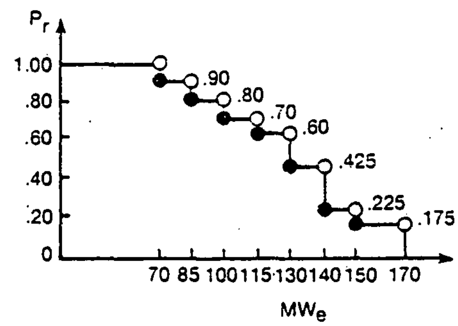


Figure 3. (a) LDC for Conventional Forecast
(b) LDC for Forecast with Uncertainty

In Table I we are given a conventional forecast; that is, hourly averages without uncertainty information for a four hour period. In Table II we are given detailed information concerning the forecast, its range at each hour, and the probability of realizing the particular values in the range.

Table I contains the information of Table II in aggregated form. Figure 3 gives load duration curves for the information in Tables I & II. We wish to note in Table II that we have chosen the distribution within an hour in unrealistic but entirely different manners. For hours 1 and 3 we have symmetric uncertainty ranges with a uniform distribution and a mean value of $\mu = 100$. For hours 2 and 4 we have an asymmetric uncertainty range with a mean value of $\mu = 150$. The fact that the distributions are not identical is extremely important for what is done below.

We would like to call attention to two things concerning the two Load Duration Curves (LDC's) of Fig. 3:

(1) the expected energy for both LDC's are equal;

(2) the second LDC is entirely consistent with the concept of an LDC which we expounded in the beginning of this section.

The fact that the expected energies are equal is immediate on an intuitive level since we are merely doing something quite similar to taking averages of averages. The mathematical justification is that we are applying the distributive law to

$$E[E(\text{forecast}, f), P] = \sum_{i=1}^{4} \sum_{j=1}^{n_i} X_{ij} f_{ij} P =$$

$$\sum_{i=1}^{4} \left(\sum_{j=1}^{n_i} X_{ij} f_{ij} P \right)$$

where

E = the expectation
f = distribution over the forecast
P = distribution over the hour
n_i = number of uncertainty points.

In this case we can consider that P equals $1/n$ identically, where n is the number of hours.

In our earlier discussion we pointed out that the ordinate values of an LDC should not be thought of as the percentage of time which the load was at value L. The interpretation we desire is that there is some chance or probability that the load will take on this value. To make this idea clear let us look at the load value of 70 MW_e. This number comes from the forecast for hours 1 and 3. The forecaster says that at those hours that load might occur with 0.20 probability. However, there is a 0.80 probability that it will not occur. When looking properly at the LDC, one must imagine oneself at some unknown instant of time in the 4 hour internal; that is, he is unaware of the specific hour. In 2 out of 4 hours he has a 0.20 probability of seeing a load of 70 MW_e. Therefore it is not correct to say that he will not see 70 MW_e 10 per cent of the time; but it is correct to say that he has a .10 probability of seeing 70 MW_e.

We have shown that the probabilistic LDC which is essential to the Baleriaux-Booth framework can handle the random variability in the forecast. Next we would like to show the effect of this uncertainty in demand. Let us continue the example by assuming generation plants with the characteristics given in Table III.

Table III. ASSUMPTIONS OF POWER PLANT CHARACTERISTICS

Machine	Nameplate Cap MW	Probability Outage	$/MWH
1	100	0.0	40
2	50	0.0	60

We have assumed 0.0 for a forced outage rate. The Baleriaux-Booth framework has gained its importance by its ability to handle forced outage in connection with load through the convolution technique. However, at this point we are concentrating on variation in load; any nonzero forced outage rate will only complicate matters and obscure the purpose of the example. Using the values in Fig. 3 and applying economic dispatch we get the following costs and reliability measure

Costs for Figure 3 (a)

L (MW)	
100	1.0 x 100 MW x $40/MWH x 4 HR = $16000
150	.5 x 50 MW x $60/MWH x 4 HR = 6000
	TOTAL COST= $22000

Probability of loss of load = 0.0

Costs for Figure 3 (b)

Cost = Production costs plus cost of expected unserved energy

L(MW)		
70	1.0	x 70 MW x $40/MWH x 4 HR = $11200
85	.9	x 15 MW x $40/MWJ x 4 HR = 2160
100	.8	x 15 MW x $40/MWH x 4 HR = 1920
115	.7	x 15 MW x $60/MWH x 4 HR = 2520
130	.6	x 15 MW x $60/MWH x 4 HR = 2160
140	.425	x 10 MW x $60/MWH x 4 HR = 1020
150	.225	x 10 MW x $60/MWH x 4 HR = 540
		TOTAL PRODUCTION COST = $21520

170 Probability of loss of load
= (0.35 x x 2) = 0.175

Expected unserved energy
= .175 x 20 MW x 4 HRS = 14 MWH

We note that the change in plant usage between Figures 3(a) and 3 (b) is not just in the high cost peaking machine (#2); (#1) also gets less usage because there is some chance that there will be less than 100 MW of demand. The case which used demand averages (a) does not recognize this possibility. We also have the peculiar situation where the costs in (b) are less than those in (a) but the reliability measures are reversed. This is explainable if one considers the extreme: if a system is 100 per cent unreliable there are no incremental fuel costs. So in (b) by fixing a cost for unserved energy greater than $60/MWH, one's intuition concerning reliability and cost is met.

We have shown that both the reliability and the production costs are dependent upon the random variability or uncertainty in the forecast, i.e., average hourly forecasts give different results than the range of values over the hour.

When one considers the position the authors have developed in previous work [1,2], one sees the implications of this result. The standard procedure for the evaluation of the worth of a solar technology is to subtract the hourly solar output from the hourly demand. The authors have shown that for wind machines and other solar sources, especially those without storage, there is a discrepancy between evaluations based on hourly inputs and those based on distributions over the hour. From the example above it appears that demand as a function of the hour and uncertainty in forecast should be expressed as

$$D = \mu_D + \tilde{\xi}_D - (\mu_S + \tilde{\xi}_S)$$

where:

- D = residual forecasted demand
- μ_D = mean of the forecast for the hour
- $\tilde{\xi}_D$ = a random variable representing the range of uncertainty in the demand
- μ_S = mean of the output from the solar source
- $\tilde{\xi}_S$ = a random variable representing the variation in output from the solar devices.

The dimensions are in MWh.

If one were to approximate the range of values in order to handle deviation values for demand for the hour, we would suggest choosing values for $\tilde{\xi}_F$, μ_S, $\tilde{\xi}_S$ as given below:

$$\tilde{\xi}_D = i\,\sigma_D \qquad i = 0,1,2,3$$

$$\tilde{\xi}_S = i/n \text{ (Range of values)} \qquad i - 1,2,\ldots,n$$

$$\mu_S = 0$$

Where σ_D is the standard deviation for the forecast for the hour and n is some reasonable number of bins or intervals for the range.

The combinatorics of the situation imply that we are taking every combination of the reasonably discretized forecast with the discretized range of output from the solar source or 7n cases for each hour.

The increased computer costs when one switches from the difference between hourly average forecast and the hourly average solar output to this combination case are not trivial. However, the pleasant surprise is that it is not necessary to do the 7n calculation for each hour; in most cases one can ignore the random variation in the forecast except at the endpoints of the LDC, the maximum and minimum demand values. The formal lemma which we prove in the Appendix states:

> Lemma: Let f(x) be a probability density defined in [a,b]. For each x in [a + ε, b − ε], (ε > 0), let there be defined one and only one density $g_x(\tau)$ τ in [x − ε, x + ε] such that x is the expected value of the distribution defined by $g_x(\tau)$.

$$\text{Define} \quad h(x) = f(x) \int_a^b g_\tau(\chi)\, d\tau.$$

> Then h(χ) if and only if $g_\chi(\chi + z) = g_{\chi - z}(\chi)$, for all χ and (z < ε). That is, the probability mass for h(χ) at each χ in [a + ε, b − ε] is the same as that for f(χ) if there exists a distribution $g_{\chi_0}(\tau)$ and all the other $g_\chi(\tau)$ are merely translated copies if that distribution.

In less formal language and in the case of a forecast, the lemma states that if the forecast range at each hour is bounded fairly tightly and if the error distribution is the same at each point in the forecast, then after one ranks the forecast values, one can forget about the variation in the forecast except perhaps at distances from the endpoints of the forecast which equal the range of uncertainty. The conclusion of course depends on the shape of the LDC.

The proof in the Appendix handles the lemma in a more formalistic approach. We would like to present at this point a more intuitive example of the Lemma. Suppose forecasts are being made at the points n = 1,2,3,4,5,6. Suppose that the mean of the forecast is the point itself, i.e., at 4, the mean of the forecast is 4. Further suppose at any interior point the forecast range only includes the adjacent points. Also at the endpoints the forecast is the point itself with probability 0.5 and the adjacent point with probability 0.5. This information is summarized in Table IV.

If one accumulates the probability masses at each point one gets the results in Table V.

From the example one gets an idea of the smoothing effect the forecast distribution has on the interior points. The uncertainty process begins to smooth out the values so they end up with the determistic mass of 1.0 and of course this is due to the uncertainty process borrowing mass from a value only to repay it from the uncertainty surrounding its neighbors.

Table IV. DATA FOR EXAMPLE BASED ON LEMMA

Point	Forecast Range	Probability
1	1	0.5
	2	0.5
2	1	0.25
	2	0.50
	3	0.25
3	2	0.25
	3	0.50
	4	0.25
4	3	0.25
	4	0.50
	5	0.25
5	4	0.25
	5	0.50
	6	0.25
6	5	0.5
	6	0.5

Table V. ACCUMULATION OF PROBABILITY MASSES IN TABLE IV

Point	Probability Mass
1	0.75
2	1.25
3	1.0
4	1.0
5	1.25
6	0.75

It is obvious that the forecasting procedure does not meet the basic assumptions of the Lemma:

(1) the range of uncertainty at any hour is unbounded.

(2) the forecast has different standard deviations (σ's) according to the following regime: when one ranks the forecasts, the extreme values have larger σ's than the interior values.

(3) $f(\chi)$ should be interpreted as the density for the occurrence of demand. In reality we are dealing with the product of probability of demand and the probability of error in forecast; that is

$$\int_a^b f(\tau)\, g_\tau(\chi)\, d\tau \quad .$$

only if $f(\tau)$ is a constant do we get

$$f(\chi) \int_a^b g_\tau(\chi)\, d\tau \quad .$$

(1) can be justified by forecasting theory. The standard assumption in forecasting is that the error of the forecast is normally distributed. Forgetting that negative forecasts are meaningless, we are left with a theoretical range of $(-\infty, \infty)$. However, the application of forecasting leans heavily on the fact that most of the probability distribution lies within 3 standard deviations of the mean; it is this fact which precludes worrying about negative forecasts. Therefore if one is to discretize the range in a rational manner, one might choose bounds of $(\mu - 3\sigma, \mu + 3\sigma)$ for the range of forecast. Since most of the probability mass is located in that region, the approximate range is both adequate and bounded.

The second problem is that there are "seams" in the forecast, areas where the distribution of uncertainty changes. In the example above we have seen the effects of changes in σ (see points 1,2,5,6 in Table V). We can justify this change in σ by examining an LDC. The slope of the LDC is much greater at the endpoints than at the interior values. This implies that there is more probability of an interior interval occuring than one of the extreme values. The obvious reason for this is that more hours have been forecasted to lie around average demands; the neighborhoods of the peak demand and, quite possibly, the minimum demands reflect very few hours and have a greater level of uncertainty. However, the lemma states that one can only dismiss the demand variability in the interior if all the uncertainty distributions are identical. It is our contention that one can make an excellent approximation by assuming constant σ except at the extremes and thereby not worry about the random variability of the forecast except in the region around the maximum and minimum values. The conventional LDC of hourly averages should only be altered then in small intervals around the endpoints. The σ's of the peak and minimum forecasts can be estimated from historical data. Using our earlier concept of a historical realization as a trajectory lying within bands around the mean value of the forecast, one can estimate these σ's. It is also quite likely that the σ for the peak forecast will be greater than that of the minimum. Some personal observations of utility data have resulted in a guess at the σ as being 6 to 9 per cent of the forecast at these points. We would then suggest an alteration of the LDC only in the neighborhood of 1 σ above and below the peak and minimum forecasts respectively. The assumption is that the σ's at other points are identical.

Finally (3) above implies that distribution over demand is uniform. That this is not so can be seen from the S shape of the LDC. However, for most LDC's a straight line can be fit through the inflection point of this S curve; that is, if one disregards intervals around the maximum and minimum points the cumulative distribution function for demand can be fit by a straight line. Therefore this interior section can be approximated by a uniform distribution and $f(\chi)$ can be removed from the integral. It is this section which we wish to approximate. The larger the range of adequate fit to the straight line, the more justification we have for ignoring the random variation in demand. This is a function of the LDC.

The conclusion on the effect of random variability of the demand is that its effects on cost are minimal but its effect on LOLP can be quite important. The increased spike in the LDC due to peak uncertainty can have a major effect. In relationship to the varibility in the output of a solar source, the random variability in demand does not dominate the picture. Rather it is the variation in output which has the major role since at this time there is not evidence which suggests that the hourly distributions over output are identical as would be required by the Lemma. If, of course, the ξ_S were shown to be identically distributed, the use of hourly averages for the output of the solar source could be justified. Finally the combinations needed to represent both random variation in the demand and variability in output are reduced by the need to examine only the endpoints of the LDC.

RESERVE MARGIN VERSUS LOLP

There are two reliability criterion used most frequently in electric utility planning: per cent reserve margin and LOLP. Per cent reserve margin is usually defined as capacity in excess of a certain percent of the forecasted peak demand. The LOLP criterion has been used in this paper and is assumed to be well-known.

Either one of these criteria make an adequate planning goal. In fact many utilities calculate an equivalency between the two criteria, recognizing the fact that the equivalency is a function of time.

There is a curious phenomenon, however; some institutions use the per cent reserve margin as the basis for planning and then make LOLP calculation separately. The reason usually expressed for this is that the per cent reserve margin takes care of any untoward contingency where the LOLP gives the loss of load risk if conditions happen as expected. The concern being expressed here may be due to the fear of the uncertainty in the forecast. If this uncertainty is restricted to random variability, the alteration of the LDC at peak to represent this variability might make the LOLP more palatable as a planning criterion. This measure would then be responsive to mix and random uncertainty.

VARIABILITY DUE TO ASSUMPTIONS

The second level of variability in demand is that due to the basic assumptions upon which the forecast is made; for what follows let us identify this level of uncertainty as scenario variation.

If we imagine a forecaster using some form of econometric forecasting tool, we can understand the demand differential as a function of the myriad of economic assumptions. We are also familiar with the current situation of different forecasts of growth rates that are filed by adversaries in various siting cases. These various growth rates would give rise to different LDC's, different production costs, and different loss of load risks.

We will show that scenario variation can also be handled in the Baleriaux-Booth framework by presenting a simplified example.

Table IV. ASSUMPTIONS FOR EXAMPLE IN VARIATION DUE TO SCENARIO

Hour	Load (MW) Scenario #1	#2	#3	Machine	Nameplate Cap (MW)	Forced Outage Rate	$/MWH
1	100	125	75	1	100	0.0	40
2	150	175	125	2	50	0.0	60

Probability for scenarios
0.2 0.5 0.3

We are given three point forecasts with no uncertainty bounds; each forecast could be assumed to represent a different rate of growth. The LDC's for the three scenarios are so trivial we will omit them. Table VII summarizes the basic statistics.

Table VII. BASIC STATISTICS OF THE THREE SCENARIOS

Scenario	Expected Energy (MWH)	Cost of Production	LOLP	Unserved Energy (MWH)
1	250	$11,000	0.0	0
2	300	$12,500	0.5	25
3	200	$ 8,500	0.0	0

Expected Values over All Scenarios

Energy = [0.2 (250 MWH) + 0.5 (300 MWH) + 0.3 (200 MWH)] = 260 MWH

Cost = 0.2 ($11,000) + 0.5 ($12,500) + 0.3 ($8,500) = $11,000

LOLP = 0.2 (0) + 0.5 (0.5) + 0.3 (0) = 0.25

Unserved Energy = 0.2 (0 MWH) + 0.5 (25 MWH) + 0.3 (0 MWH) = 12.5 MWH

To show the simplicity of the example we give the calculation of costs for scenario #1:

1.0 x 100 MW x $40/MWH x 2 HRS = $ 8,000

0.5 x 50 MW x $60/MWH x 2 HRS = $ 3,000

$11,000

Now if one regards the three scenarios as forecasts of the amount of demand and the probability of demand for each hour, one gets the following Table.

Table VIII. (a) SCENARIOS TREATED AS FORECASTS (b) PR [LOAD > L] FOR FORECASTS

HR	(a) Demand (MW)	Probability	(b) L (MW)	Pr [Load > L]
	75	0.3	0	1.00
1	100	0.2	75	0.85
	125	0.5	100	0.75
	125	0.3	125	0.35
2	150	0.2	150	0.25
	175	0.5	175	0.0

In the (a) section of Table VIII we have handled the scenarios just as we did the forecasts earlier, i.e., the values given are the range for the hour. In the (b) part of the table we have weighted the probability by the hourly weight of 1/2 and formed the values for the LDC table given below

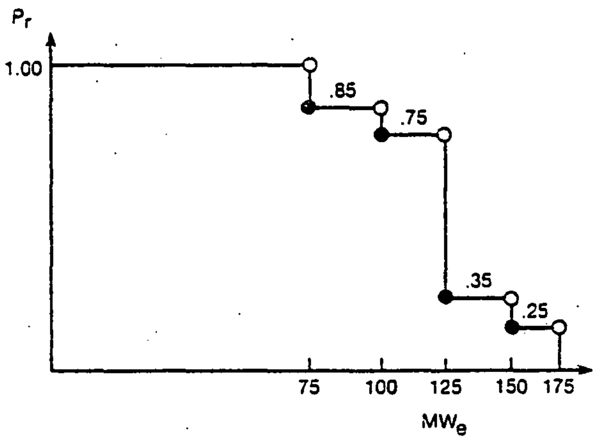


Figure 4. LDC for Scenarios Treated as Forecasts

From the LDC and the machine characteristics we get the following statistics:

Expected Energy =
(1.0 x 75 MW) + (0.85 x 25 MW) + (0.75 x 25 MW) + (0.35 x 25 MW) + (0.25 x 25 MW) x 2 HRS
= 260 MWH

Cost =	1.0	x 75 MW	x $40/MWH	x 2 HRS	=	$ 6,000
	0.85	x 25 MW	x $40/MWH	x 2 HRS	=	$ 1,700
	0.75	x 25 MW	x $60/MWH	x 2 HRS	=	$ 2,250
	0.35	x 25 MW	x $60/MWH	x 2 HRS	=	$ 1,050
						$11,000

LOLP = 0.25

Expected Unserved Energy = 0.25 x 25 MW x 2 HR
= 12.5 MWH

These statistics are as expected: from the expected values over the scenarios one gets the same results as taking the distribution over the demand and calculating the expected values. We also note that if one wanted to take into account random variability of forecast given a particular scenario, we would proceed as before. The Baleriaux-Booth framework handles the use of scenario variability as well as random variability. It requires only that one view the LDC as giving probabilities with respect to an instant of time.

While the Baleriaux-Booth framework handles scenario variability, it is questionable whether there is any value in using the technique for this kind of uncertainty. The expected values for cost and reliability are pertinent for random variation. Given a set of assumptions, costs and service failure due to randomness are conditions of life. They can be considered unavoidable risks. However, costs and risks due to scenario construction are a different matter. The individual costs and LOLP for each scenario is important. The planner is concerned with the risks of a planning schedule in the face of demand uncertainties. When one takes expected values over all the scenarios as one does in treating the scenarios as forecasts, one loses the individual results from the scenario. They become aggregated and smoothed by the expectation process. For the planner it's quite important, as shown in our example, that he realize that he may be facing costs of $12500 and 25 MWH of unserved energy from Scenario #2, the most probable scenario. The smoothing that occurs in the respective expected values of $11000 and 12.5 MWH could be misleading information.

With regards to the relationship between variation in the output from the intermittent source and that from scenario variation, it is usually stated that large differences in rates of demand growth will dominate if the pentration of the intermittent sources is small in comparison to the growth rates. But this is basically a misleading statement. If one choses to combine all scenarios with appropriate weights as we have done above, scenario variation dominates. However, if the scenarios are placed individually into the Baleriaux-Booth framework then, as was shown in the random variation section of this paper, the variation in the intermittent output is the important concept. If one is interested in evaluating risks in the scenarios, it is important to be concerned about the variability in output.

CONCLUSIONS

The major result of the paper is that if one is concerned with a forecast, which has been ranked or ordered so that chronological order is lost, random variation in the forecast can be ignored except in the neighborhood of this endpoints if the error distributions are identical at every point in the forecast. If it were true that output from intermittent sources were identically distributed and if one subtracted the output from the load on an hourly basis and then ranked the residuals, variation in output could be ignored. However, it is more reasonable to assume that the variation in output of intermittent sources will be a function of the mean of the hourly output; and since the mean will vary diurnally for most intermittent sources, the hourly distributions will not be identical. Therefore the variation of output must be considered. The random variation of the forecast must only be considered near the peak and minimum demands. However, variation in demand due to assumptions can have a major effect on costs and reliability. Unless one is willing to lose the individual results from a scenario through the smoothing effects of expectation, the effect of output variation should still be considered on a scenario by scenario basis.

APPENDIX

Lemma: Let $f(x)$ be a probability density defined on $[a,b]$. For each x in $[a + \varepsilon, b - \varepsilon]$,

($\varepsilon > 0$), let there be defined one and only one density $g_\chi(\tau)$, τ in $[x - \varepsilon, x + \varepsilon]$ such that x is the expected value of the distribution defined by $g_\chi(\tau)$. Define:

$$h(x) = f(x) \int_a^b g_\tau(\chi)\, d\tau$$

Then $h(\chi) = f(\chi)$ if and only if $g_\chi(\chi + z) = g_{\chi - z}(\chi)$, for all χ and ($z < \varepsilon$). That is, the probability mass for $h(\chi)$ at each χ in $[a + \varepsilon, b - \varepsilon]$ is the same as that for $f(\chi)$ if there exists a distribution $g_{\chi_0}(\tau)$ and all the other $g_\chi(\tau)$ are merely translated copies of that distribution.

[Before giving the proof we would like to make a few clarifications on the assumptions of the lemma for the continuous case. For each point in the subinterval we are defining secondary distributions as in the case of forecasting a μ and acknowledging an uncertainty around this μ. Therefore at each x, probability mass is being accumulated from the densities which have expected values in the neighborhood of x. For this given x the accumulation has a magnitude of 1 if the secondary densities (error distributions) are identical except for translation. We also note that only expected values in the $[x - \varepsilon, x + \varepsilon]$ neighborhood of x can contribute probability mass to x].

Proof: we first prove the <u>if</u> part.

Since $g_\chi(\tau)$ is a density defined on $[x - \varepsilon, \chi + \varepsilon]$, for all x in $[a + \varepsilon, b - \varepsilon]$

$$g_\chi(\tau) = 0 \qquad \tau \notin (x - \varepsilon, x + \varepsilon)$$

$$\int_{x-\varepsilon}^{x+\varepsilon} g_\chi(\tau)\, d\tau = 1.$$

Then for all x in $[a + \varepsilon, b - \varepsilon]$

$$h(x) = f(x) \int_a^b g_\tau(\tau)\, d\tau = f(x) \int_{x-\varepsilon}^{x+\varepsilon} g_\tau(\chi)\, d\tau$$

But

$$g_{\chi-z}(\chi) = g_\chi(\chi+z) \text{ for all z such that } 0<z<\varepsilon$$

and

$$g_{\chi-z}(\chi) = g_\chi(\chi+z) \text{ for all z such that } -\varepsilon<z<0 .$$

Therefore

$$h(x) = f(x) \int_{x-\varepsilon}^{x+\varepsilon} g_\tau(\chi)\, d\tau$$

$$= f(x) \int_{\chi-\varepsilon}^{\chi+\varepsilon} g_\chi(\tau) d\tau = f(\chi)$$

For the <u>only if</u> part we will manufacture a counterexample, showing that if $g_\chi(\tau)$ is not identical the result does not hold. The counterexample will be for a discrete distribution without any loss of generality since integrals and summations could be interchanged in the above or alternatively the integrals could be interpreted as Riemann-Stieltjes integrals.

Let $f(n)$ be a uniform distribution on $n = 1,2,\ldots,10$, that is $f(n) = 1/10$ $n = 1,2,\ldots,10$

Let $\varepsilon = 1$

Let $g_n(\tau)$ be defined in the following manner:

for $n = 1$	$g_1(1) = 1$
for $n = 10$	$g_{10}(10) = 1$
for $n = 5$	$g_5(4) = 0.1$ $g_5(5) = 0.8$ $g_5(6) = 0.1$
for all other n	$g_n(n - 1) = 0.25$ $g_n(n) = 0.50$ $g_n(n + 1) = 0.25$

Since $\varepsilon = 1$ our concern is for the points $2,3,4,\ldots,9$. We have defined $g_n(\tau)$ identically except for $n = 5$.

For 2,3,7,8,9

$$h(x) = f(x)$$

as can be seen in the calculation of h(2):

$$h(2) = f(2) \sum_{i=1}^{3} g_i(2) = f(2)\ (0.25+0.50+0.25)$$
$$= f(2)$$

But for 4,5,6

$$h(x) \neq f(x).$$

For $n = 4$

$$h(4) = f(4) \sum_{i=3}^{5} g_i(4) = f(4)\ [.025+0.5+0.1]$$
$$= 0.85\ f(4)$$

$$h(5) = f(5) \sum_{i=4}^{6} g_i(4) = f(5)\ [0.25+0.8+0.25]$$
$$= 1.3\ f(5)$$

$$h(6) = f(6) \sum_{i=5}^{7} g_i(6) = f(6)\ [0.1+0.5+0.25]$$
$$= 0.85\ f(6)$$

[In intuitive terms it is at the seams, the points where the distributions differ, that probability mass starts to accumulate to values other than 1].

ACKNOWLEDGEMENT

The authors wish to thank the Wind Systems Branch of DOE whose encouragement and funding were essential to the work.

REFERENCES

[1] Fegan, G. R.; Percival, C. D. "Integration of Intermittent Sources into Baleriaux-Booth Production Cost Models" presented at IEEE Winter Power Meeting, February 1980, in New York, N.Y. A 80 111-5.

[2] Fegan, G. R.; Percival, C. D. "Problems with the Integration of Intermittent Sources into Utility Evaluation Models" to be presented at the International Symposium in Circuits and Systems, April 1980, in Houston, Texas.

[3] Booth, R. R. "Generation Planning Considering Uncertainty." Proceedings of the 1971 IEEE PICA Conference, 71-C26-PWR, pp. 62-66.

[4] Booth, R. R. "The A-B-C of Probabilistic Simulation." Unpublished paper.

[5] Booth, R. R. "Power System Simulation Model Based on Probability Analysis." Proceedings of the IEEE PICA Conference, 71-C26-PWR, pp. 285-291.

[6] Baleriaux, H.; Jamulle, Fr. Linard de Guertechin. "Simulation de l'exploitation d'un pace de machines thermiques de production d'electricite' couple a des stations de parapages." Review E (edition SRBE) pp. 3-24, vol V, No. 7, 1967.

George R. Fegan possesses masters degrees in American Literature and mathematics and a Ph.D in Mathematical Statistics. The latter degree was received from Oregon State University. Dr. Fegan has taught mathematics and statistics at universities and colleges in Oregon and California. He has been a principal partner in a consulting firm which did statistical research for B.P.A., electric utilities, and Oregon Medical School. From 1977 through 1979 he was a member of the Corporate Planning Division of Portland General Electric; in April of 1979 he joined the Utility Application Branch of the Solar Energy Research Institute.

Dr. Fegan's research area is stochastic processes. He has published papers in the areas of utility planning, calculation of radioactive decay and transmutation, and reactor safety.

C. David Percival received his B.S. and M.S. degrees from Trinity University, San Antonio, Texas, in 1971 and 1979, respectively in Engineering Science.

From 1971 to 1978 he was with the San Antonio City Public Service Board. His last three years were in the Generation Planning and Fuels section. He joined the Solar Energy Research Institute in 1978 as a staff engineer in the Systems Analysis Branch. His current responsibilities include the analysis of solar applications to electric utility generating systems.